Interior Point Algorithms II

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye
Typically, a single function driven algorithm is more preferred to take possible large step sizes, rather than check and balance on multiple measures.

For $\mathbf{x} \in \text{int} \ F_p$ and $(\mathbf{y}, \mathbf{s}) \in \text{int} \ F_d$, the joint primal-dual potential function is defined by

$$
\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n + \rho) \log(\mathbf{x}^T \mathbf{s}) - \sum_{j=1}^{n} \log(x_j s_j),
$$

where $\rho \geq 0$.

Then, for $\rho > 0$, $\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \to -\infty$ implies that $\mathbf{x}^T \mathbf{s} \to 0$. More precisely, we have

$$
\mathbf{x}^T \mathbf{s} \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) - n \log n}{\rho}\right).
$$
Once have a pair \((x^k, y^k, s^k) \in \text{int } F\), we compute direction vectors \(d_x\), \(d_y\) and \(d_s\) from the system equations:

\[
S^k d_x + X^k d_s = \frac{(x^k)^T s^k}{n+\rho} e - X^k S^k e, \\
A d_x = 0, \\
-A^T d_y - d_s = 0.
\] (1)

Note that \(d_x^T d_s = -d_x^T A^T d_y = 0\) here. Then choose a step-size scalar \(\theta(>0)\) and assign

\[
x^{k+1} = x^k + \theta d_x > 0, \quad y^{k+1} = y^k + \theta d_y, \quad s^{k+1} = s^k + \theta d_s > 0.
\]

This is the Newton method for the optimality conditions/equations of the potential minimization problem:

\[
XSe = \frac{(x^k)^T s^k}{n+\rho} e, \\
A x = b, \\
-A^T y - s = -c.
\] (2)
To simplify rotations, let

\[ \mathbf{d}_{x'} + \mathbf{d}_{s'} = \mathbf{r}' := (X S)^{-0.5}(X^T S^{0.5} e - X S e), \]

\[ A' \mathbf{d}_{x'} = 0, \]

\[ -(A')^T \mathbf{d}_y - \mathbf{d}_{s'} = 0. \]

where

\[ D = X^{0.5} S^{-0.5}, \quad A' = AD, \quad \mathbf{d}_{x'} = D^{-1} \mathbf{d}_x, \quad \mathbf{d}_{s'} = D \mathbf{d}_s. \]

Again, we maintain \( \mathbf{d}_{x'}^T \mathbf{d}_{s'} = 0. \)

Unlike in the path-following algorithm, \( \|\mathbf{r}'\|^2 \) may be too big to make \( \mathbf{x} + \mathbf{d}_x \) or \( \mathbf{s} + \mathbf{d}_s \) positive. So that we need to add a step size \( \theta \) to scale \( \mathbf{r}' \) such that it makes new iterate feasible.
Lemma 1 Let the direction vector \( \mathbf{d} = (d_x, d_y, d_s) \) be generated by equation (2), and let

\[
\theta = \frac{\alpha \sqrt{\min(XSe)}}{\|r'\|},
\]

where \( \alpha \) is a positive constant less than 1. Let

\[
x^+ = x + \theta d_x, \quad y^+ = y + \theta d_y, \quad \text{and} \quad s^+ = s + \theta d_s.
\]

Then, we have \((x^+, y^+, s^+) \in \text{int} \ F\) and

\[
\psi_{n+\rho}(x^+, s^+) - \psi_{n+\rho}(x, s) 
\leq -\alpha \sqrt{\min(XSe)} \| (XS)^{-1/2} (e - \frac{(n + \rho)}{x^T s} Xs) \| + \frac{\alpha^2}{2(1 - \alpha)}.
\]
We first present a technical lemma:

**Lemma 2** If \( \mathbf{d} \in \mathbb{R}^n \) such that \( \|\mathbf{d}\|_\infty < 1 \) then

\[
e^T \mathbf{d} \geq \sum_{i=1}^{n} \log(1 + d_i) \geq e^T \mathbf{d} - \frac{\|\mathbf{d}\|^2}{2(1 - \|\mathbf{d}\|_\infty)}.
\]

The proof is based on the Taylor expansion of \( \ln(1 + d_i) \) for \(-1 < d_i < 1\).
Figure 1: Logarithmic approximation by linear and quadratic functions
Proof Sketch of the Theorem

It is clear that \( A\mathbf{x}^+ = \mathbf{b} \) and \( A^T\mathbf{y}^+ + \mathbf{s}^+ = \mathbf{c} \). We now show that \( \mathbf{x}^+ > 0 \) and \( \mathbf{s}^+ > 0 \). This is similar to the previous proof for the path-following algorithm

\[
\|\theta X^{-1}\mathbf{d}_x\|_2^2 + \|\theta S^{-1}\mathbf{d}_s\|_2^2 \leq \theta^2 \frac{\|\mathbf{r}'\|^2}{\min(\mathbf{XSe})} = \frac{\alpha^2 \min(\mathbf{XSe})}{\|\mathbf{r}'\|^2} \frac{\|\mathbf{r}'\|^2}{\min(\mathbf{XSe})} = \alpha^2 < 1.
\]

Therefore,

\[
\mathbf{x}^+ = \mathbf{x} + \theta \mathbf{d}_x = X(e - \theta X^{-1}\mathbf{d}_x) > 0
\]

and

\[
\mathbf{s}^+ = \mathbf{s} + \theta \mathbf{d}_s = S(e - \theta S^{-1}\mathbf{d}_s) > 0.
\]
Sketch of the proof continued

\[
\psi(x^+, s^+) - \psi(x, s) \\
= (n + \rho) \log \left(1 + \frac{\theta d_s^T x + \theta d_x^T s}{x^T s}\right) - \sum_{j=1}^{n} \left(\log(1 + \frac{\theta d_s^j}{s^j}) + \log(1 + \frac{\theta d_x^j}{x^j})\right) \\
\leq (n + \rho) \left(\frac{\theta d_s^T x + \theta d_x^T s}{x^T s}\right) - \sum_{j=1}^{n} \left(\log(1 + \frac{\theta d_s^j}{s^j}) + \log(1 + \frac{\theta d_x^j}{x^j})\right) \\
\leq \frac{n + \rho}{x^T s} \theta(d_s^T x + d_x^T s) - \theta e^T (S^{-1} d_s + X^{-1} d_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= \theta \left(\frac{n + \rho}{x^T s} e^T (X d_s + S d_x) - e^T (S^{-1} d_s + X^{-1} d_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= \theta \left(\frac{n + \rho}{x^T s} e^T (X d_s + S d_x) - e^T (XS)^{-1} (X d_s + S d_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= \theta \left(\frac{n + \rho}{x^T s} XSe - e\right)^T (XS)^{-1} (X d_s + S d_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= \theta \left(\frac{n + \rho}{x^T s} XSe - e\right)^T (XS)^{-1} \left(\frac{x^T s}{n + \rho} e - XSe\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= -\theta \cdot \frac{n + \rho}{x^T s} \cdot \|r'\|^2 + \frac{\alpha^2}{2(1-\alpha)} = -\alpha \sqrt{\min(XSe)} \cdot \frac{n + \rho}{x^T s} \cdot \|r'\| + \frac{\alpha^2}{2(1-\alpha)}. \]
Let $v = XSe$. Then, we can prove the following technical lemma:

**Lemma 3** Let $v \in \mathbb{R}^n$ be a positive vector and $\rho \geq \sqrt{n}$. Then,

$$\sqrt{\min(v)}\|V^{-1/2}(e - \frac{(n + \rho)}{e^Tv}v)\| \geq \sqrt{3/4}.$$ 

Combining these two lemmas we have

$$\psi_{n+\rho}(x^+, s^+) - \psi_{n+\rho}(x, s)$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1 - \alpha)} = -\delta$$

for a constant $\delta$. 
Description of Algorithm

Given \((x^0, y^0, s^0) \in \text{int } \mathcal{F}\). Set \(\rho \geq \sqrt{n}\) and \(k := 0\).

While \((x^k)^T s^k \geq \epsilon\) do

1. Set \((x, s) = (x^k, s^k)\) and \(\gamma = n/(n + \rho)\) and compute \((d_x, d_y, d_s)\) from (2).

2. Let \(x^{k+1} = x^k + \bar{\alpha} d_x, y^{k+1} = y^k + \bar{\alpha} d_y\), and \(s^{k+1} = s^k + \bar{\alpha} d_s\) where

\[
\bar{\alpha} = \arg \min_{\alpha \geq 0} \psi_{n+\rho}(x^k + \alpha d_x, s^k + \alpha d_s).
\]

3. Let \(k := k + 1\) and return to Step 1.
Theorem 1 Let $\rho \geq \sqrt{n}$ and $\psi_{n+\rho}(x^0, s^0) \leq \rho \log((x^0)^T s^0) + n \log n$. Then, the Algorithm terminates in at most $O(\rho \log((x^0)^T s^0 / \epsilon))$ iterations with

$$\langle x^k, s^k \rangle = c^T x^k - b^T y^k \leq \epsilon.$$  

$$\begin{align*}
\langle x^k, s^k \rangle &\leq \exp\left(\frac{\psi_{n+\rho}(x^0, s^0) - n \log n}{\rho}\right) \\
&\leq \exp\left(\frac{\psi_{n+\rho}(x^0, s^0) - n \log n - \rho \log((x^0)^T s^0 / \epsilon)}{\rho}\right) \\
&\leq \exp\left(\frac{\rho \log(x^0, s^0) - \rho \log((x^0)^T s^0 / \epsilon)}{\rho}\right) \\
&= \exp(\log(\epsilon)) = \epsilon.
\end{align*}$$

The adaptively search of best $\rho$?
Termination with Exact Optimizers

- The first is a “cross-over” procedure to find a basic feasible solution (BFS, corner point) whose objective value is at least as good as the current interior point. Let $A$, $b$, $c$ be integers and $L$ be their bit length, and let a second best BFS solution be $x^{2nd}$ and the optimal objective value be $z^*$. Then

$$c^T x^{2nd} - z^* > 2^{-L}.$$ 

Thus, one can terminate interior-point algorithm when

$$c^T x^k - b^T y^k \leq 2^{-L}.$$ 

- The second approach is to compute a strictly complementary solution pair. The method uses the primal-dual interior-point pair to identify the strict complementarity partition $(P^*, Z^*)$ and then “purify or project” the primal interior solution onto the primal optimal face and the dual interior solution onto the dual optimal face, based on the following theorem:

**Theorem 2** Given an interior solution $x^k$ and $s^k$ in the solution sequence generated by an
interior-point algorithm, define

\[ P^k = \{ j : x^k_j \geq s^k_j, \ \forall j \} \quad \text{and} \quad Z^k = \{ 1, ..., n \} \setminus P^k. \]

Then, we have \( P^k = P^* \) whenever

\[ c^T x^k - b^T y^k \leq 2^{-L}. \]

Thus, the worst-case iteration bound for interior-point algorithms is \( O(\sqrt{nL}) \) if the initial point pair \( (x^0)^T s^0 \leq 2^L. \)
Combining the primal and dual into a single linear feasibility problem, then applying LP algorithms to find a feasible point of the problem. Theoretically, this approach can retain the currently best complexity result.

- The big $M$ method, i.e., add one or more artificial column(s) and/or row(s) and a huge penalty parameter $M$ to force solutions to become feasible during the algorithm.

- Phase I-then-Phase II method, i.e., first try to find a feasible point (and possibly one for the dual problem), and then start to look for an optimal solution if the problem is feasible and bounded.

- Combined Phase I-Phase II method, i.e., approach feasibility and optimality simultaneously. To our knowledge, the “best” complexity of this approach is $O(n \log(R/\epsilon))$. 
Homogeneous and Self-Dual Algorithm

- It solves the linear programming problem without any regularity assumption concerning the existence of optimal, feasible, or interior feasible solutions, while it retains the currently best complexity result.

- It can start at any positive primal-dual pair, feasible or infeasible, near the central ray of the positive orthant (cone), and it does not use any big $M$ penalty parameter or lower bound.

- Each iteration solves a system of linear equations whose dimension is almost the same as that solved in the standard (primal-dual) interior-point algorithms.

- If the LP problem has a solution, the algorithm generates a sequence that approaches feasibility and optimality simultaneously; if the problem is infeasible or unbounded, the algorithm will produce an infeasibility certificate for at least one of the primal and dual problems.
A pair of LP has two alternatives

(Solvable) \[ Ax - b = 0 \]
\[-A^T y + c \geq 0, \]
\[ b^T y - c^T x = 0, \]
\[ y \text{ free}, \ x \geq 0 \]

(Infeasible) \[ Ax = 0 \]
\[-A^T y \geq 0, \]
\[ b^T y - c^T x > 0, \]
\[ y \text{ free}, \ x \geq 0 \]
An Integrated Homogeneous System

The two alternative systems can be homogenized as one:

\[
(HP) \quad A x - b \tau = 0 \\
- A^T y + c \tau = s \geq 0, \\
b^T y - c^T x = \kappa \geq 0, \\
y \text{ free, } (x; \tau) \geq 0
\]

where the two alternatives are

(Solvable) : \((\tau > 0, \kappa = 0)\) or (Infeasible) : \((\tau = 0, \kappa > 0)\)
The Homogeneous System is Self-Dual

\((HP)\) \quad Ax - b\tau = 0, \quad (y')

\((-A^T y + c\tau = s \geq 0, \quad (x')\)

\((HD)\) \quad Ax' - b\tau' = 0,

\((-A^T y' - c\tau' \leq 0, \quad (x')\)

\(b^T y - c^T x = \kappa \geq 0, \quad (\tau')\)

\(y \) free, \quad (x; \tau) \geq 0

\(x^T s + \tau \kappa = 0.\)

**Theorem 3** System \((HP)\) is feasible (e.g. all zeros) and any feasible solution \((y, x, \tau, s, \kappa)\) is self-complementary:

Furthermore, it has a strictly self-complementary feasible solution

\[
\begin{pmatrix}
    x + s \\
    \tau + \kappa
\end{pmatrix}
\geq 0,
\]
Let’s Find Such a Feasible Solution

Given $x^0 = e > 0$, $s^0 = e > 0$, and $y^0 = 0$, we formulate

$$(HSDP) \quad \min \quad \theta$$

s.t. $\begin{align*}
Ax - b\tau + \bar{b}\theta &= 0, \\
-A^Ty + c\tau - \bar{c}\theta &\geq 0, \\
b^Ty - c^Tx + \bar{z}\theta &\geq 0,
\end{align*}$$

$y$ free, $x \geq 0$, $\tau \geq 0$, $\theta$ free,

where

$$\begin{align*}
\bar{b} &= b - Ae, \\
\bar{c} &= c - e, \\
\bar{z} &= c^Te + 1.
\end{align*}$$

But it may just give us the all-zero solution.
Let's try to add one more constraint to prevent the all-zero solution

\[(HSDP) \quad \min \quad (n + 1)\theta\]

s.t. \quad \begin{align*}
    &Ax - b\tau + \bar{b}\theta = 0, \\
    &-A^T y + c\tau - \bar{c}\theta \geq 0, \\
    &b^T y - c^T x + \bar{z}\theta \geq 0, \\
    &-\bar{b}^T y + \bar{c}^T x - \bar{z}\tau = -(n + 1), \\
    &y \text{ free, } x \geq 0, \quad \tau \geq 0, \quad \theta \text{ free.}
\end{align*}\]

Note that the constraints of \((HSDP)\) form a skew-symmetric system and the objective coefficient vector is the negative of the right-hand-side vector, so that it remains a self-dual linear program.

\((y = 0, \ x = e, \ \tau = 1, \ \theta = 1)\) is a strictly feasible point for \((HSDP)\).
\[(HSDP) \quad \min \quad (n + 1)\theta\]
\[
\text{s.t.} \quad Ax - b\tau + \bar{b}\theta = 0, \\
-A^T y + c\tau - \bar{c}\theta = s \geq 0, \\
b^T y - c^T x + \bar{z}\theta = \kappa \geq 0, \\
-\bar{b}^T y + \bar{c}^T x - \bar{z}\tau = -(n + 1), \\
y \text{ free, } \quad x \geq 0, \quad \tau \geq 0, \quad \theta \text{ free.}
\]

Denote by \(\mathcal{F}_h\) the set of all points \((y, x, \tau, \theta, s, \kappa)\) that are feasible for (HSDP). Denote by \(\mathcal{F}_h^0\) the set of interior feasible points with \((x, \tau, s, \kappa) > 0\) in \(\mathcal{F}_h\). By combining the constraints, we can derive the last (equality) constraint as
\[
e^T x + e^T s + \tau + \kappa - (n + 1)\theta = (n + 1),
\]
which serves indeed as a normalizing constraint for (HSDP) to prevent the all-zero solution.
Theorem 4  Consider problems (HSDP) and (HSDD).

i) (HSDD) has the same form as (HSDP), i.e., (HSDD) is simply (HSDP) with \((y, x, \tau, \theta)\) being replaced by \((y', x', \tau', \theta')\).

ii) (HSDP) has a strictly feasible point

\[ y = 0, \quad x = e > 0, \quad \tau = 1, \quad \theta = 1, \quad s = e > 0, \quad \kappa = 1. \]

iii) (HSDP) has an optimal solution and its optimal solution set is bounded.

iv) The optimal value of (HSDP) is zero, and

\[(y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h \quad \text{implies that} \quad (n + 1)\theta = x^T s + \tau\kappa.\]

v) There is an optimal solution \((y^*, x^*, \tau^*, \theta^* = 0, s^*, \kappa^*) \in \mathcal{F}_h\) such that

\[
\begin{pmatrix}
   x^* + s^* \\
   \tau^* + \kappa^*
\end{pmatrix} > 0,
\]
which we call a strictly self-complementary solution. (Similarly, we sometimes call an optimal solution to (HSDP) a self-complementary solution; the strict inequalities above need not hold.)
Theorem 5  Let $\left( y^*, x^*, \tau^*, \theta^* = 0, s^*, \kappa^* \right)$ be a strictly self complementary solution for (HSDP).

i) (LP) has a solution (feasible and bounded) if and only if $\tau^* > 0$. In this case, $x^*/\tau^*$ is an optimal solution for (LP) and $\left( y^*/\tau^*, s^*/\tau^* \right)$ is an optimal solution for (LD).

ii) (LP) has no solution if and only if $\kappa^* > 0$. In this case, $x^*/\kappa^*$ or $s^*/\kappa^*$ or both are certificates for proving infeasibility: if $c^T x^* < 0$ then (LD) is infeasible; if $-b^T y^* < 0$ then (LP) is infeasible; and if both $c^T x^* < 0$ and $-b^T y^* < 0$ then both (LP) and (LD) are infeasible.
**Theorem 6** i) For any $\mu > 0$, there is a unique $(y, x, \tau, \theta, s, \kappa)$ in $\mathcal{F}_h^0$, such that

\[
\begin{pmatrix}
Xs \\
\tau \kappa
\end{pmatrix} = \mu e.
\]

ii) Let $(d_y, d_x, d_\tau, d_\theta, d_s, d_\kappa)$ be in the null space of the constraint matrix of (HSDP) after adding surplus variables $s$ and $\kappa$, i.e.,

\[
\begin{align*}
Ad_x - bd_\tau + \bar{b}d_\theta &= 0, \\
-A^T d_y + cd_\tau - \bar{c}d_\theta - d_s &= 0, \\
b^T d_y - c^T d_x + zd_\theta - d_\kappa &= 0, \\
-\bar{b}^T d_y + \bar{c}^T d_x - \bar{z}d_\tau &= 0.
\end{align*}
\]

\[(d_x)^T d_s + d_\tau d_\kappa = 0.\]
Endogenous Potential Function and Central Path

\[ \psi_{n+\rho}(x, s, \tau, \kappa) := (n + 1 + \rho) \log(x^T s + \tau \kappa) - \sum_{j=1}^{n} \log(x_j s_j) - \log(\tau \kappa), \]

and

\[ \mathcal{C} = \left\{ (y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h^0 : \begin{pmatrix} X s \\ \tau \kappa \end{pmatrix} = \frac{x^T s + \tau \kappa}{n + 1} \mathbf{e} \right\}. \]

Obviously, the initial interior feasible point proposed in Theorem 4 is on the path with \( \mu = 1 \) or \( (x^0)^T s^0 + \tau^0 \kappa^0 = n + 1 \).
Consider solving the following system of linear equations for \((d_y, d_x, d_\tau, d_\theta, d_s, d_\kappa)\) that satisfies (4) and
\[
\begin{pmatrix}
X d_s + S d_x \\
\tau^k d_\kappa + \kappa^k d_\tau
\end{pmatrix} = \gamma \mu e - \begin{pmatrix} X s \\ \tau \kappa \end{pmatrix}.
\]

**Theorem 7** The \(O(\sqrt{n} \log((x^0)^T s^0 / \epsilon))\) interior-point algorithm, coupled with a termination technique described above, generates a strictly self-complementary solution for (HSDP) in \(O(\sqrt{n} (\log(c(A, b, c)) + \log n))\) iterations and \(O(n^3 (\log(c(A, b, c)) + \log n))\) operations, where \(c(A, b, c)\) is a positive number depending on the data \((A, b, c)\). If (LP) and (LD) have integer data with bit length \(L\), then by the construction, the data of (HSDP) remains integral and its length is \(O(L)\). Moreover, \(c(A, b, c) \leq 2^L\). Thus, the algorithm terminates in \(O(\sqrt{n} L)\) iterations and \(O(n^3 L)\) operations.
Example

Consider the example where

\[
A = \begin{pmatrix} -1 & 0 & 0 \end{pmatrix}, \quad b = 1, \quad \text{and} \quad c = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}.
\]

Then,

\[
y^* = 2, \quad x^* = (0, 2, 1)^T, \quad \tau^* = 0, \quad \theta^* = 0, \quad s^* = (2, 0, 0)^T, \quad \kappa^* = 1
\]

could be a strictly self-complementary solution generated for (HSDP) with

\[
c^T x^* = 1 > 0, \quad by^* = 2 > 0.
\]

Thus \((y^*, s^*)\) demonstrates the infeasibility of (LP), but \(x^*\) doesn’t show the infeasibility of (LD). Of course, if the algorithm generates instead \(x^* = (0, 1, 2)^T\), then we get demonstrated infeasibility of both.
Software Implementation

Cplex, GUROBI

SEDUMI: http://sedumi.mcmaster.ca/

MOSEK: http://www.mosek.com/products_mosek.html

IPOPT: https://projects.coin-or.org/Ipopt

hsdLPsolver: Sparse Linear Programming Solver (Matlabe .m file).